FIFTY YEARS OF X-RAY DIFFRACTION

DEDICATED TO THE
INTERNATIONAL UNION OF CRYSTALLOGRAPHY
ON THE OCCASION OF THE
COMMEMORATION MEETING IN MUNICH
JULY 1962

by

P. P. EWALD

EDITOR

AND NUMEROUS CRYSTALLOGRAPHERS

PUBLISHED FOR THE
INTERNATIONAL UNION OF CRYSTALLOGRAPHY
BY N.V. A. OOSTHOEK'S UITGEVERSMAATSCHAPPIJ
UTRECHT, THE NETHERLANDS

Contents

CHAPTER 13

Problems of Biochemical Structures

by Ralph W. G. Wyckoff

The distinction between organic and biological chemistry becomes progressively less sharp as chemists learn to establish the molecular configurations and to synthesize more and more of the constituents of living matter. During the past generation such complicated natural products as sterols, alkaloids and antibiotics have been synthesized, as well as many derivatives which nature has not had occasion to make. From the standpoint of X-ray diffraction analysis the line between the organic and the biological has also become increasingly blurred as our more powerful tools have been able to establish molecular structures too difficult for the unaided chemistry of today. Nevertheless it remains worthwhile to consider as a group apart those native products we have not yet learned to prepare in the laboratory; and this is the field to be treated in the present chapter.

The extreme complexity of living matter made it inevitable that knowledge of what it is should lag far behind our understanding of the inanimate world. Biochemistry, and now our X-ray techniques are giving a rapidly increasing insight into the molecular composition of plants and animals—but this is not in itself enough. In all but the simplest manifestations of life, elaborateness of organization is as striking as complexity of chemical composition, and knowledge of this organization is an essential part of any understanding we may acquire of the mechanism of life. It is a basic problem of the biophysics now emerging to ascertain the details of this organization, which often involves an order in molecular arrangement that X-ray diffraction can effectively interpret. We can no longer summarize in a few pages what is now known of this natural order; instead we shall seek to review the growth of our appreciation of the value of X-ray methods and the appearance of other methods that have come to supplement the information X-rays can give.

Biology in its development over the last century has shown how cells

are organized into tissues and how in the higher animals these tissues are associated into the organs whose corporate life becomes that of living individuals. Until the advent of X-ray diffraction it was generally considered that, except for teeth and bones, these ingredients of living matter were amorphous, gel-like colloids and that consequently structures showing order in arrangement were scarcely to be expected at and below the cellular level. Discovery of how rare was the amorphous condition even among colloids was one of the important early consequences of X-ray diffraction.

As a result, many of the fibrous colloids of biology were examined during the 1920's and found to give diffraction effects pointing to a partial crystallinity or a definite but sub-crystalline order in particle arrangement. This was true of such widely diverse substances as animal hair and birds' feathers, tendons and other connective tissues of animals, and cellulosic fibers from all sorts of plants. In general the diffraction phenomena were too fragmentary to provide a satisfactory picture of the kinds of order responsible for them but they were the basis for a continuing study of these natural products. It was during this period that Herzog fostered the application of X-ray methods in his institute for cellulose research, that Sponsler suggested that cellulose was a parallel packing of polysaccharide crystallites and that somewhat later Meyer and Mark gave a first crystalline interpretation to the X-ray pattern of cellulose. At this time, too, Astbury began the studies centered around hair and wool which, proceeding uninterruptedly over the next thirty years, have contributed so much of the foundation upon which an intimate knowledge of protein structure is now being built. Katz' observation that rubber yields an oriented crystalline diffraction pattern only after stretching demonstrated how X-rays could reveal reversible molecular rearrangements; this experiment was a prototype for many of the most rewarding subsequent studies of fibrous structures.

Such exploratory examinations of natural fibers were actively continued and broadened throughout the 1930's. Especially significant for the future were Astbury's demonstration of the alpha, beta and super-contracted states of the keratin-like fibrous proteins and the way they pass from one to another. He pictured this in terms of the coiling of peptide chains which forecast the spiral models we have now come to consider as underlying all protein structures.

The work of this period also developed a clear distinction between fibrous and globular proteins and the different types of molecules responsible for each. As they occur in nature the fibrous proteins are

insoluble whereas the proteins we designate as globular function in solution. The latter can often be crystallized in the laboratory and these crystals have the same regularities in external form and the same optical characteristics as the crystals of simple substances. The ultracentrifugal investigations initiated by Svedberg in the middle 1920's determined their molecular weights and led to the discovery that their enormous molecules were more or less globular in shape and had the same uniformity in size that prevails among small molecules. When the still larger virus proteins came to be recognized the ultracentrifuge demonstrated their molecular uniformity and supplied a new way to purify them as well as other substances too unstable to be handled without damage by the existing chemical procedures.

Several attempts to obtain diffraction patterns from crystals of globular proteins were made early in the 1930's. At first these met with mediocre success, largely because of the enormous amounts of water in the crystals studied and the ease with which this water and the crystalline structure are lost. With time, however, conditions were established for the preparation of photographs from substantially undamaged single crystals. Their patterns contained thousands of reflections and were evidently caused by the same kind of three-dimensional regularity that prevails in chemically simple crystals. They indicated the experimental basis for an ultimate determination of atomic arrangement while at the same time emphasizing the magnitude of the task. Perutz began his lifelong study of crystalline hemoglobin at this time. The success now being achieved is overwhelmingly due to his skill and perseverance in what must often have seemed a hopeless task, as well as to W. L. Bragg's scientific and material support over the years. This success is a triumph of persistence and faith which is doubly noteworthy now when research projects which do not offer an assured and quick result are rarely able to survive even when they can be begun.

Though the fibrous proteins dealt with during the pre-war years did not yield the many X-ray data obtainable from globular proteins, they did provide much knowledge upon which striking post-war advances have been based. Chemical work suggested that most fibrous solids could not be put into solution and then reconstituted the way a globular protein can be repeatedly dissolved and recrystallized; and ultracentrifugation of solutions obtained from wool, cellulose and other fibrous materials revealed non-uniform particles that could best be interpreted as molecular fragments. X-ray experiments made at this time on reconstituted collagen were chiefly important in showing that

such was not necessarily the case. Precipitates from dissolved tendon gave the same X-ray patterns and thus were essentially the same as the native product. The preparation of molecularly ordered structures in the laboratory as suggested by these and analogous experiments with cellulose has since led to a wide extension of our knowledge of fibrous substances.

After the war the X-ray study of the biologically important proteins experienced the same stimulus as did other phases of crystallography. There have been a number of reasons for this, not the least of which has been a more general appreciation of the importance of crystallography itself. In its wake has come increased financial support for the many new workers who have turned to this field. Other factors that have contributed to the current rapid growth of our knowledge of biological structures have been the introduction of improved techniques of X-ray experimentation, of computers able to deal with the great masses of data these improved methods have made possible and of several new and complementary sources of information such as chromatographic amino acid analysis and electron microscopy. These methods have contributed very differently to the understanding we have of the structures of various biological solids and it is important to see what these contributions have been if we are to appreciate how our understanding has expanded. The order we find in such solids is of different kinds and our methods are not equally effective in dealing with them. There is either (1) a true crystallinity which expresses itself as a three-dimensional regularity in molecular arrangement, (2) a paracrystallinity consisting of one- or two-dimensionally repeated molecular packings such as are shown by the fibrous proteins or (3) repetitive atomic distributions within individual molecules. In general X-ray diffraction is most effective where the order is greatest while the electron microscope, for instance, has been most helpful where the ordering is less than crystalline. Thus it is the electron microscope rather than X-ray diffraction which has revealed the laminations of the lipoidal sheats of myelinated nerves and the chloroplasts of plants. The contributions of X-ray diffraction and of electron microscopy have been more equal for the nucleic acids. Neither they nor the nucleoproteins are known to furnish naturally occurring ordered solids but in the laboratory nucleic acids sometimes give X-ray fibre diagrams analogous to those obtained from the fibrous proteins. Both the ultracentrifuge and the electron microscope show these acids and proteins to have thread-like molecules of indefinite lengths and we may expect their solids to be parallel aggregates of these filaments. Chemical analysis

shows that all nucleic acids consist of linear associations of a few nucleotides somewhat as the polypeptide chains of proteins are strings of amino acid residues. Following the success of Pauling and Corey's helical structure for proteins, Watson and Crick have proposed an analogous linking of nucleotides as the basis for nucleic acids; and such a chain is compatible with the limited X-ray data existing preparations can supply. The visibility of individual filamentous molecules of nucleic acids under the electron microscope makes it certain that combined X-ray and microscopic studies will be increasingly fruitful as we learn to grow specimens with better molecular ordering.

Cellulose is in some respects a most promising and in other respects one of the most disappointing of natural fibers. Its promise lies in the richness of the X-ray data it often yields. It is disappointing in the sense that we are more dependent on natural products for experimental material than is the case with some of the fibrous proteins; the crystallinity of derivatives made in the laboratory is not yet sufficient to provide all the data required for a thoroughgoing determination of structure. Natural cellulose, however, illustrates in an interesting fashion the different kinds of order that have been mentioned. Its elementary particles as seen under the electron microscope are indefinitely long fibrils that commonly have a diameter of ca. 200 Å. Each of these is made up, in part at least, of minute crystallites whose atomic arrangement is responsible for the sharp X-ray reflections we observe. Since the original proposal of structure by Meyer and Mark, these reflections have been made the experimental basis for atomic arrangements assigned to a number of forms and derivatives of cellulose but we need many more data to make these determinations complete. In the secondary walls of plants the elementary fibrils are in parallel arrays. Some information about them was early gained from X-ray fiber diagrams but by portraying each particle separately, the electron microscope has superceded diffraction as the tool for investigating such arrays. It not only has shown the fibrillar arrangement but has told much about how cellular activity creates and orients the fibrils themselves.

Most has been learned about the order that prevails in protein solids. Though we can sometimes see the molecular arrangement in protein crystals with the electron microscope, the developing power of X-ray methods to establish complete atomic arrangements makes them preeminent. With the paracrystalline fibrous proteins the X-ray data are less complete while the electron microscope has more to show about their molecular architecture.

The early X-ray work of Astbury and others resulted in a natural classification of the fibrous proteins. To one group belong the keratinous proteins such as those which constitute the skin and horny structures of mammals, the feathers of birds, the scales of reptiles and the flagella of some bacteria. Mammalian keratin normally gives what Astbury called the alpha-keratin X-ray pattern. These keratins elongate on stretching and then give the different, beta-keratin, pattern which is also produced by avian and reptilian keratin. Beta-keratin will return to its original alpha form when tension is released (under suitable conditions) and all keratin can be made to shrink to a supercontracted state still more shortened than the alpha. Recently there has been wide interest in showing that this behaviour of keratin and related fibrous structures is compatible with the helical polypeptide model of Pauling and Corey. With the electron microscope one can see the elementary fibrils of such keratinous solids as wool, but the imperfect order of their arrangement automatically restricts the information this instrument can give.

The collagenous proteins are very different from keratin both in X-ray fiber diagram and in physical and chemical characteristics. They are not, like keratin, extensible and they yield solutions from which solids showing high degrees of molecular ordering can be precipitated. Sometimes this order is the same as that observed in nature, under other conditions it is equally good but different. X-ray diffraction has supplied information about this order as it exists in tendon and also about repetitions within the separate fibrils but during the post-war years the application of the electron microscope has contributed far more to our knowledge of both the intra- and inter-fibrillar order of these several polymorphous forms of collagen. An interpretation of what has been learned is now becoming feasible with the help of knowledge of the amino acid sequences in the polypeptide chains from which the molecular filaments are built up. The chromatographic analysis has shown that along the polypeptide chains of collagen every third residue is glycine and that hydroxyproline is its second most numerous amino acid constituent; and we can in a general way picture how reactions involving mainly this residue could result in the several ordered molecular associations we observe. Making use of these amino acid distributions a molecular structure involving three intertwined Pauling and Corey polypeptide helices has recently been proposed which is compatible with what is now known about collagen. Such semi-speculative molecular models as this, like the analogous Watson and Crick model for nucleic acid, are of great assistance in

thinking about these exceedingly complex ingredients of living matter and in correlating their properties. They are not firmly established by existing data but the cumulative weight of the comparisons to be made between such models and experiment will with time give a clearer indication of their validity even though the prospect of obtaining enough X-ray data for complete structure determinations is at present very poor.

Muscle is a fibrous tissue whose order, even more elaborate than that of collagen, cannot be inferred from its X-ray diffractions but is being revealed in detail by the electron microscope. More than one kind of filamentous molecule can thus be seen in intact muscle, distributed regularly with respect to one another in an almost crystalline array. Muscular contraction patently involves interaction between these ordered elements of an especially complex 'multicomponent' crystalline structure, and electron microscopy has been giving mounting insight into how muscles work. As would be expected from the different sized molecular filaments seen in muscle, several fibrous proteins can be obtained from it; and after separation from one another they furnish ordered, almost crystalline solids whose study is usefully supplementing what can be seen in muscle itself.

The intimate knowledge now being acquired of the crystal structures of hemoglobin and myoglobin and the corresponding advances being made with other globular proteins, notably ribonuclease, is, as already indicated, the culmination of a generation's development of X-ray methods and the first step towards an entirely new understanding of proteins as the substrate of all life. To appreciate the significance of these determinations of structure we need to take note of the stages through which their analysis has proceeded and the new techniques that have made each advance possible. The pre-war studies of hemoglobin demonstrated its essential crystallinity and the availability of the experimental evidence required for a deeper penetration into its structure. The very richness of these data, however, brought new and formidable problems. Thus the precise measurement and identification of the more than 20000 reflections that can be obtained from the crystals of a protein required X-ray equipment that did not exist before the war and demanded a mass of routine measurement never before encountered in the course of crystallographic work. Methods of manual computation were completely unable to deal with the data thus produced and accordingly advance in this field could only follow the development and availability of high speed computers.

The early determination of the dimensions of the unit cell set certain

limitations on the shape of the hemoglobin and myoglobin molecules but progress towards a direct deduction of structure hinged on being able to find the correct signs for measured structure factors. This involved using the heavy metal isomorphous replacement technique that had already led to complete structures for several complicated organic crystals. It required in this case the introduction into the protein molecule of mercury or other heavy atoms without damage to the molecule as a whole or a change in the molecular distribution within the crystal. Much difficult chemical work was necessary to accomplish this but the success that has been achieved in making these metal substitutions augures well for future advances involving other crystalline proteins. In the case of hemoglobin and myoglobin, analysis of the data from the native protein and its derivatives led to knowledge of an increasing number of signs and thus to Fourier summations that pointed to important atomic concentrations within the volumes occupied by individual molecules. The more detailed step of passing from such general, preliminary ideas of molecular shape to a picture of the way the various polypeptide chains are distributed and interconnected has been made by taking advantage of two other sources of information. One of these, to which we have already referred, was complete chemical analysis to establish the amino acid residues present in these substances and their sequence in the polypeptide chains; the other was use of the alpha helix of Pauling and Corey to assign configurations to these chains. In this way probable structures could be limited and direct attacks continued. The interplay of these procedures has led to structures first for the chemically simpler myoglobin by Kendrew and his co-workers and now for hemoglobin. Though they do not yet define the exact positions of all the thousands of atoms in a unit cell they already give very probable pictures of the spatial distribution of the various amino acid residues in the molecules and undoubtedly soon will define these atomic distributions more closely. Probably before long we will have a correspondingly detailed knowledge of ribonuclease and perhaps of other globular protein crystals.

The scientific importance of these results for an understanding of proteins is obvious but they are equally important as an incontrovertible justification of the large amounts of work and money they have entailed. They make it clear that other proteins can be similarly analyzed and at a fraction of the cost in time and money that have gone into these first structures. Such protein analyses will always require programmes having assured support for a group of workers over a

number of years but we may now be sure that, given this support, the job can be done. It seems inevitable that the analyses of myoglobin and hemoglobin are but the first steps in an intensive study of protein structure which can scarcely fail to advance in unpredictable and truly revolutionary ways our knowledge of the mechanism of life.